(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 November 2004 (25.11.2004)

PCT

(10) International Publication Number
WO 2004/101781 A1

(51) International Patent Classification[7]: C12N 9/50, A61P 31/14, G01N 33/50

(21) International Application Number: PCT/EP2004/005109

(22) International Filing Date: 13 May 2004 (13.05.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/469,818 13 May 2003 (13.05.2003) US

(71) Applicant *(for all designated States except US)*: **UNIVERSITÄT ZU LÜBECK** [DE/DE]; Ratzeburger Allee 160, 23538 Lübeck (DE).

(72) Inventors; and

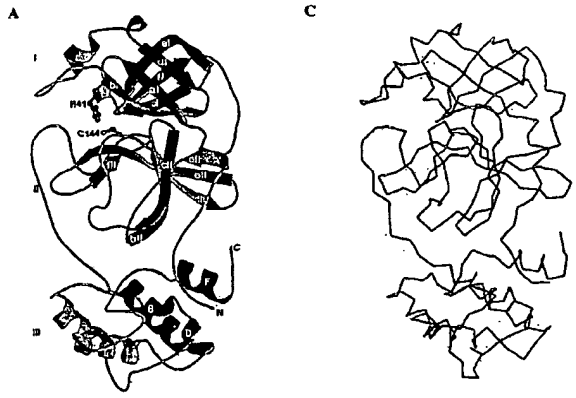
(75) Inventors/Applicants *(for US only)*: **HILGENFELD, Rolf** [DE/DE]; Arnimstrasse 28b, 23566 Lübeck (DE). **ANAND, Kanchan** [—/DE]; Peter-Monnik-Weg 4, 23538 Lübeck (DE). **ZIEBUHR, John** [DE/DE]; Haendelstr. 14, 97209 Veitschoechheim (DE). **MESTERS, Jeroen, R.** [DE/DE]; Eichhörnchenweg 8, 23627 Gross-Grönau (DE). **WADHWANI, Parvesh** [—/DE]; Vogelsang 27, 76229 Karlsruhe (DE).

(74) Agent: **BIEHL, Christian**; Boehmert & Boehmert, Niemannsweg 133, 24105 Kiel (DE).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

[Continued on next page]

(54) Title: CRYSTAL STRUCTURE OF HUMAN CORONAVIRUS 229E MAIN PROTEINASE, AND USES THEREOF FOR DEVELOPING SARS INHIBITORS



(57) Abstract: The invention relates to SARS inhibitors. Human coronaviruses are major causes of upper respiratory tract illness in humans, in particular, the common cold. Recent investigations have shown that a novel coronavirus causes the Severe Acute Respiratory Syndrome (SARS), a disease that is characterized by high fever, malaise, rigor, headache, non-productive cough or dyspnea and which is rapidly spreading. Within the scope of the invention, based on the structural analysis of the binding mode of the SARS M^{pro} enzyme a group of prototype inhibitors is provided that acts as suitable drugs targeting a majority of viral infections of the respiratory tract, including SARS.

WO 2004/101781 A1

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designation US*

Published:

— *with international search report*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

CRYSTAL STRUCTURE OF HUMAN CORONAVIRUS 229E MAIN PROTEINASE, AND USES THEREOF FOR DEVELOPING SARS INHIBITORS

FIELD AND BACKGROUND OF THE INVENTION

Human coronaviruses (HCoV) are major causes of upper respiratory tract illness in humans, in particular, the common cold *(1)*. To date, only the 229E strain of HCoV has been characterized in detail because it used to be the only isolate that grows efficiently in cell culture. It has recently been shown that a novel coronavirus causes the Severe Acute Respiratory Syndrome (SARS), a disease that is rapidly spreading from its likely origin in Southern China to several countries in other parts of the world *(2,3)*. SARS is characterized by high fever, malaise, rigor, headache, non-productive cough or dyspnea and may progress to generalized, interstitial inflitrates in the lung, requiring incubation and mechanical ventilation *(4)*. The fatility rate among persons with illness meeting the current definition of SARS is around 15% (calculated on outcome, i.e. deaths/(deaths + recovered patients)). Epidemiological evidence suggests that the transmission of this newly emerging pathogen occurs mainly by face-to-face contact, although airborne transmission cannot be fully excluded. By May 05, 2003, more than 6400 cases of SARS had been diagnosed world-wide, with the numbers still rapidly increasing. At present, no efficacious therapy is available.

Coronaviruses are positive-stranded RNA viruses featuring the largest viral RNA genomes known to date (27-31 kb). The human coronavirus 229E replicase gene, encompassing more than 20,000 nucleotides, encodes two overlapping polyproteins, pp1a (≈450 kD) and pp1ab (≈750 kD) *(5)* that mediate all the functions

required for viral replication and transcription *(6)*. Expression of the COOH-proximal portion of pp1ab requires (-1) ribosomal frameshifting *(5)*. The functional polypeptides are released from the polyproteins by extensive proteolytic processing. This is primarily achieved by the 33.1-kDa HCoV main proteinase (M^{pro}) *(7)*, also called 3C-like proteinase or $3CL^{pro}$, which cleaves the polyprotein at 11 conserved sites involving mostly Leu-Gln↓(Ser,Ala,Gly) sequences, a process initiated by the enzyme's own autolytic cleavage from pp1a and pp1ab *(8,9)*. The functional importance of M^{pro} in the viral life cycle makes this proteinase an attractive target for the development of drugs directed against SARS and other coronavirus infections.

The design of anticoronaviral drugs directed against the viral main proteinases requires the availability of data on the three-dimensional structures of the target enzymes. In 2002, we determined the crystal structure of the M^{pro} of transmissible gastroenteritis virus (TGEV), a coronavirus infecting pigs *(10)*. The structure revealed that coronavirus M^{pro} consists of three domains, the first two of which together distantly resemble chymotrypsin. However, the catalytic site comprises a Cys-His dyad rather than the Ser-His-Asp triad found in typical chymotrypsin-like serine proteinases.

SUMMARY OF THE INVENTION

We determined the crystal structure, at 2.6 Å resolution, of the free enzyme of human coronavirus (strain 229E) M^{pro} ***(claim A1,* PDB file no. 1)**. Further, we constructed a three-dimensional model for the M^{pro} of SARS coronavirus (SARS-CoV) ***(claim A2,* PDB file no. 2)**, based on our crystal structures for HCoV and TGEV M^{pro}s (*claim A1* and *(10)*) and on the genomic sequence of SARS-CoV *(11)*. SARS-CoV M^{pro} shares 40 and 44% amino-acid sequence identity with its

TGEV and HCoV counterparts, respectively. We also analyzed the putative cleavage sites of M^{pro} in the viral polyprotein as derived from the genomic sequence *(11)* and found them to be highly similar to those of M^{pro}s of HCoV, TGEV and other coronaviruses. Further, we developed a method to produce recombinant SARS-CoV M^{pro} and modifications (mutants) thereof ***(claim B)***. We show that the recombinant wild-type enzyme exhibits proteolytic activity while an active-site mutant (C145A) does not. We demonstrate that recombinant SARS-CoV M^{pro} cleaves a pentadecapeptide representing the NH_2-terminal autocleavage site of TGEV main proteinase. Comparison of the crystal structures for HCoV and TGEV M^{pro} and the model for SARS-CoV M^{pro} shows that the substrate-binding sites are well conserved among coronavirus main proteinases.

In order to determine the exact binding mode of the substrate and to enable the structure-based design of drugs directed at coronavirus M^{pro}, we have synthesized the substrate-analog chloromethyl ketone inhibitor Cbz-Val-Asn-Ser-Thr-Leu-Gln-CMK, the sequence of which was derived from the P4 - P1 residues of the NH_2-terminal autoprocessing site of HCoV M^{pro}. We have determined the 2.37 Å crystal structure of a complex between this inhibitor and porcine transmissible gastroenteritis (corona)virus (TGEV) main proteinase (***claim A3,* PDB file no. 3**). Analysis of the binding mode of this inhibitor shows that it is similar to that seen for an inhibitor of the distantly related human rhinovirus 3C proteinase *(12)*. On the basis of the combined structural information, a group of prototype inhibitors, **1**, is proposed that should block all these enzymes and thus be suitable drugs targeting a majority of viral infections of the respiratory tract, including SARS (***claim C***).

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1: Three-dimensional structure of coronavirus M^{pro}. **A.** ***(illustrating claim A1;*** **PDB file no. 1):** Monomer of HCoV M^{pro}. Domains I (top), II, and III (bottom) are indicated. Helices are red and strands green. α-helices are labeled A to F according to occurence along the primary structure, with the additional one-turn A' α-helix in the N-terminal segment (residues 11 - 14). β-strands are labeled a to f, followed by an indication of the domain to which they belong (I or II). NH_2- and COOH-terminus are labeled N and C, respectively. Residues of the catalytic dyad, Cys^{144} and His^{41}, are indicated. **B.** ***(illustrating claims A1, A2):*** Structure-based sequence alignment of the main proteinases of coronaviruses from all three groups. HCoV, human coronavirus 229E (group I); TGEV, porcine transmissible gastroenteritis virus (group I); MHV, mouse hepatitis virus (group II); BCoV, bovine coronavirus (group II); SCoV, SARS coronavirus (between groups II and III); IBV, avian infectious bronchitis virus (group III). The autocleavage sites of the proteinases are marked by vertical arrows above the sequences. In addition to the sequences of the mature enzymes, four residues each of the viral polyprotein NH_2-terminal to the first and COOH-terminal to the second autocleavage site are shown. Note the conservation of the cleavage pattern, (small)-Xaa-Leu-Gln↓(Ala,Ser,Gly). Thick bars above the sequences indicate α-helices (numbered A', A to F); horizontal arrows indicate β-strands (numbered a-f, followed by the domain to which they belong). Residue numbers for HCoV M^{pro} are given below the sequence; 3-digit numbers are centered about the residue labeled. Symbols in the second row below the alignment mark residues involved in dimerization of HCoV and TGEV M^{pro}: open circle (o), only main chain involved; asterisk (*), only side chain involved; plus (+), both main chain and side chain involved. From the almost absolute conservation of side chains involved in dimerization, it can be concluded that SARS-CoV M^{pro} also has the capacity to form

dimers. In addition, *side chains* involved in inhibitor binding in the TGEV M^{pro} complex are indicated by triangles (Δ), and catalytic-site residues Cys^{144} and His^{41} as well the conserved "$Y^{160}MH^{162}$" motif are shaded. **C:** ***(illustrating claim A2;* PDB file no. 2):** Cα plot of a monomer of M^{pro} as model-built on the basis of the crystal structures of HCoV 229E M^{pro} and TGEV M^{pro}. Residues identical in HCoV M^{pro} and M^{pro} are indicated in red.

Figure 2 ***(illustrating claims A1, A2;* PDB file no. 1):** Dimer of HCoV M^{pro}. The NH_2-terminal residues of each chain squeeze between domains II and III of the parent monomer and domain II of the other monomer. NH_2- and COOH-termini are labeled by cyan and magenta spheres, and letters N and C, respectively.

Figure 3. A ***(illustrating claim A3;* PDB file no. 3):** Refined model of the TGEV M^{pro}-bound hexapeptidyl chloromethyl ketone inhibitor built into electron density (2||Fo|-|Fc||, contoured at 1 σ above the mean). There was no density for the Cbz group and for the Cβ atom of the P1 Gln. Inhibitor shown in red, protein in gray; Cys^{144} is yellow. **B:** Inhibitors will bind to different coronavirus M^{pro}s in an identical manner. Superimposition (stereo image) of the substrate-binding regions of the free enzymes of HCoV 229E M^{pro} (blue; PDB file no. 1) and SARS-CoV M^{pro} (magenta; PDB file no. 2), and of TGEV M^{pro} (green; PDB file no. 3) in complex with the hexapeptidyl chloromethyl ketone inhibitor (red; PDB file no. 3). The covalent bond between the inhibitor and Cys^{144} of TGEV M^{pro} is in orange.

Figure 4 ***(illustrating claim B):*** A TGEV M^{pro} cleavage site is recognized and cleaved by recombinant SARS-CoV M^{pro}. The peptide H_2N-VSVNSTLQ↓SGLRKMA-COOH (vertical arrow indicates the cleavage site),

representing the NH_2-terminal autoprocessing site of TGEV M^{pro}, was efficiently cleaved by M^{pro} but not by an inactive catalytic-site mutant of this enzyme. HPLC elution profiles of **A**, uncleaved peptide (incubated with buffer) in the absence of proteinase; **B**, peptide incubated with M^{pro}; **C**, peptide incubated with M^{pro}-C145A.

Figure 5 ***(illustrating claim C)*****:** Derivatives of the antirhinoviral drug AG7088 should inhibit coronavirus M^{pro}s. Superimposition (stereo image) of the substrate-binding regions of TGEV M^{pro} (green) in complex with the hexapeptidyl chloromethyl ketone inhibitor (red) and HRV2 $3C^{pro}$ (marine) in complex with the inhibitor AG7088 (yellow).

Figure 6 ***(illustrating claim C)*****:** Derivatives of AG7088, compounds **1**, proposed for inhibition of coronavirus main proteinases, including SARS coronavirus (SARS-CoV) M^{pro}. P2 = *p*-fluoro-benzyl: AG7088. We claim all derivatives of this compound, with any P2 group. We would also like to claim more distantly related compounds, such as AA1-AA2-AA3-AA4-P2-Gln-vinylogous ester (also the methyl and isopropylester, and other alkyl), with AA1, AA2, AA4: any amino acid or absent; AA3: small (such as Thr, Val, Ser, Ala); P2: Leu, Phe, Met, and derivatives thereof.

MAKING AND USING THE INVENTION

Claim A1: The crystal structure of HCoV M^{pro} shows that the molecule comprises three domains (Fig. 1A). Domains I and II (residues 8-99 and 100-183, respectively) are six-stranded antiparallel β-barrels and together resemble the architecture of chymotrypsin and of picornavirus 3C proteinases. The substrate-binding site is located in a cleft between these two domains. A long loop (residues 184 to 199) connects domain II to the COOH-terminal domain (domain III, residues 200-300). This latter domain, a globular cluster of five helices, has been implicated in the proteolytic activity of M^{pro} *(13)*. The HCoV M^{pro} structure is very similar to that of TGEV M^{pro} *(10)*. The r.m.s. deviation between the two structures is ~1.5 Å for all 300 Cα positions of the molecule* but the isolated domains exhibit r.m.s. deviations of only ~0.8 Å. With HCoV 229E and TGEV both being group I coronaviruses *(14)*, their main proteinases share 61% sequence identity.

**Footnote:* The construct of HCoV M^{pro} used in this work lacks two amino acid residues from the COOH-terminus. HCoV M^{pro} Δ(301-302) has the same enzymatic properties as full-length HCoV M^{pro} but yields much superior crystals. In the structure of full-length M^{pro}, residues 301 and 302 are disordered and not seen in the electron density.

Claim A2: For comparison of its enzymatic properties with those of the HCoV and TGEV M^{pro}s, we have expressed SARS-CoV (strain TOR2) M^{pro} in *E. coli*** and preliminarily characterized the proteinase. The amino-acid sequence of SARS-CoV M^{pro} displays 40 and 44% sequence identity to HCoV 229E M^{pro} and TGEV M^{pro}, respectively (see Fig. 1B for a structure-based alignment). Identity levels are 50% and 49%, respectively, between SARS-CoV M^{pro} and the corresponding proteinases

from the group II coronaviruses, mouse hepatitis virus (MHV) and bovine coronavirus (BCoV). Finally, M^{pro} shares 39% sequence identity with avian infectious bronchitis virus (IBV) M^{pro}, the only group III coronavirus for which a main proteinase sequence is available. These data are in agreement with the conclusion deducible from the sequence of the whole genome *(11)* that the new virus is most similar to group II coronaviruses, although some common features with IBV (group III) can also be detected.

***Footnote:* SARS-CoV M^{pro} from strain TOR2; acc: AY274119, SARS-CoV pp1a/pp1ab residues 3241 to 3544

The level of similarity between SARS-CoV M^{pro} and HCoV as well as TGEV M^{pro}s allowed us to construct a reliable three-dimensional model for SARS-CoV M^{pro} (Fig. 1C). There are three 1- or 2-residue insertions in M^{pro}, relative to the structural templates; as to be expected, these are all located in loops and do not present a problem in model building. Interestingly, domains I and II show a higher degree of sequence conservation (42-48% identity) than domain III (36%-40%) between SARS-CoV M^{pro} and the coronavirus group I enzymes.

Claims A1 and A2: HCoV 229E M^{pro} forms a tight dimer (contact interface, predominantly between domain II of molecule A and the NH_2-terminal residues of molecule B: ~1300 $Å^2$) in the crystal, with the two molecules oriented perpendicular to one another (Fig. 2). Our previous crystal structure of the TGEV M^{pro} *(10)* revealed the same type of dimer. We could show by dynamic light scattering that both HCoV and TGEV M^{pro} exist as a mixture of monomers (~65%) and dimers (~35%) in diluted solutions (1-2 mg proteinase/ml). However, since the architecture of the dimers

including most details of intermolecular interaction are the same in both TGEV M^{pro} (three independent dimers per asymmetric unit) and HCoV M^{pro} (one dimer per asymmetric unit), *i.e.*, in completely different crystalline environments, we believe that dimer formation is of biological relevance in these enzymes. In the M^{pro} dimer, the NH_2-terminal amino-acid residues are squeezed in between domains II and III of the parent monomer and domain II of the other monomer, where they make a number of very specific interactions that appear tailor-made to bind this segment with high affinity after autocleavage. This mechanism would immediately enable the catalytic site to act on other cleavage sites in the polyprotein. However, the exact placement of the amino terminus also seems to have a structural role for the mature M^{pro}, since deletion of residues 1 to 5 lead to a decrease in activity to 0.3% in the standard peptide-substrate assay *(10)*. Nearly all side chains of TGEV M^{pro} and HCoV M^{pro} involved in formation of this dimer (marked in Fig. 1B) are conserved in the SARS-CoV enzyme so that it is safe to assume a dimerization capacity for the latter as well.

Claims A1, A2: In the active site of HCoV M^{pro}, Cys^{144} and His^{41} form a catalytic dyad. In contrast to serine proteinases and other cysteine proteinases, which have a catalytic triad, there is no third catalytic residue present. HCoV M^{pro} has Val^{84} in the corresponding position (Cys in SARS-CoV M^{pro}), with its side chain pointing away from the active site. A buried water molecule is found in the place that would normally be occupied by the third member of the triad; this water is hydrogen-bonded to His^{41} Nδ1, Gln^{163} Nε2, and Asp^{186} Oδ1 (His, His, and Asp in both SARS-CoV and TGEV M^{pro}).

Claim A3: To allow structure-based design of drugs directed at coronavirus M^{pro}s, we sought to determine the exact binding mode of M^{pro} substrates. To this end, we synthesized the substrate-analog chloromethyl ketone inhibitor Cbz-Val-Asn-Ser-Thr-Leu-Gln-CMK ('CMK' in what follows) and soaked it into crystals of TGEV M^{pro} because these were of better quality and diffracted to higher resolution than those of HCoV M^{pro}. The sequence of the inhibitor was derived from the P6 - P1 residues of the NH_2-terminal autoprocessing site of TGEV M^{pro} (SARS-CoV M^{pro} and HCoV M^{pro} have Thr-Ser-Ala-Val-Leu-Gln and Tyr-Gly-Ser-Thr-Leu-Gln, respectively, at the corresponding positions; see Fig. 1B). X-ray crystallographic analysis at 2.37 Å resolution revealed difference density for all residues (except the benzyloxycarbonyl (Cbz) protective group) of the inhibitor, in two (B and F) out of the six TGEV M^{pro} monomers in the asymmetric unit (Fig. 3A). In these monomers, there is a covalent bond between the Sγ atom of Cys^{144} and the methylene group of the chloromethyl ketone.

There are no significant differences between the structures of the enzyme in the free and in the complexed state. The substrate-analog inhibitor binds in the shallow substrate-binding site at the surface of the proteinase, between domains I and II (Fig. 3A). The residues Val-Asn-Ser-Thr-Leu-Gln occupy, and thereby define, the subsites S6 to S1 of the proteinase. Residues P5 to P3 form an antiparallel β-sheet with segment 164-167 of the long strand eII on one side, and they also interact with segment 189-191 of the loop linking domains II and III on the other (Fig. 3A). The functional significance of this latter interaction is supported by the complete loss of proteolytic activity upon deletion of the loop region in TGEV M^{pro} *(10).*

In coronavirus M^{pro} polyprotein cleavage sites, the P1 position is invariably occupied by Gln. At the very bottom of the M^{pro} S1 subsite, the imidazole of His^{162} is

suitably positioned to interact with the P1 glutamine side chain (Figs. 3A,B). The required neutral state of His^{162} over a broad pH range appears to be maintained by two important interactions: *i)*, stacking onto the phenyl ring of Phe^{139}, and *ii)*, accepting a hydrogen bond from the hydroxyl group of the buried Tyr^{160}. In agreement with this structural interpretation, any replacement of His^{162} completely abolishes the proteolytic activity of HCoV and feline coronavirus (FIPV) M^{pro} *(13,15)*. Furthermore, FIPV M^{pro} Tyr^{160} mutants have their proteolytic activity reduced by a factor of >30 *(15)*. All of these residues are conserved in M^{pro} and, in fact, in all coronavirus main proteinases. Other elements involved in the S1 pocket of the M^{pro} are the main-chain atoms of Ile^{51}, Leu^{164}, Glu^{165}, and His^{171}. In M^{pro}, Ile^{51} becomes Pro and Leu^{164} is Met, although this is less relevant since these residues contribute to the subsite with their main-chain atoms only (Fig. 3B; side chains involved in specificity pockets are marked by "Δ" in Fig. 1B).

Apart from a few exceptions, coronavirus M^{pro} cleavage sites have a Leu residue in the P2 position *(8)*. The hydrophobic S2 subsite of the proteinase is formed by the side chains of Leu^{164}, Ile^{51}, Thr^{47}, His^{41} and Tyr^{53}. The corresponding residues in SARS-CoV M^{pro} are Met, Pro, Asp, His and Tyr. In addition, residues 186 - 188 line the S2 subsite with some of their main-chain atoms. The Leu side chain of the inhibitor is well accommodated in this pocket. It is noteworthy that M^{pro} has an alanine residue (Ala^{46}) inserted in the loop between His^{41} and Ile^{51}, but this is easily accommodated in the structural model and does not change the size or chemical properties of the S2 specificity site (see Fig. 3B).

There is no specificity for any particular side chain at the P3 position of coronavirus M^{pro} cleavage sites. This agrees with the P3 side chain of our substrate analog being oriented towards bulk solvent. At the P4 position, there has to be a small amino-acid residue such as Ser, Thr, Val, or Pro because of the congested

cavity formed by the side chains of Leu^{164}, Leu^{166}, and Gln^{191} as well as the main-chain atoms of Ser^{189}. These are conserved or conservatively substituted (L164M, S189T) in SARS-CoV M^{pro}. The P5 Asn side chain interacts with the main chain at Gly^{167}, Ser^{189}, and Gln^{191} (Pro, Thr, Gln in the enzyme), thus involving the loop linking domains II and III, whereas the P6 Val residue is not in contact with the protein. Although the inhibitor used in the present study does not include a P1' residue, it is easily seen that the common small P1' residues (Ser, Ala, or Gly) can be easily accommodated in the S1' subsite of TGEV M^{pro} formed by Leu^{27}, His^{41}, and Thr^{47}, with the latter two residues also being involved in the S2 subsite (Leu, His, and Asp in M^{pro}). Superimposition of the structures of the TGEV M^{pro}-CMK complex and the free enzyme of HCoV M^{pro} shows that the two substrate-binding sites are basically the same (Fig. 3B). All residues along the P site of the cleft are identical, with the exception of the conservative M190L replacement (Ala in SARS-CoV M^{pro}). In other coronavirus species including the SARS pathogen, M^{pro} residues 167 and 187 - 189 show some substitutions but since these residues contribute to substrate binding with their main-chain atoms only, the identity of the side chains is less important. Indeed, the substrate-binding site of the SARS-CoV M^{pro} model matches those of its TGEV and HCoV counterparts perfectly (Fig. 3B). Thus, there is no doubt that the CMK inhibitor will bind to the HCoV M^{pro} and SARS-CoV M^{pro} as well as all other coronavirus homologs with similar affinity and in the same way as it does to TGEV M^{pro}.

Claim B: We developed a method to express SARS-CoV M^{pro} in *E. coli*, as a fusion protein with maltose-binding protein (MBP). The free SARS-CoV M^{pro} was released from this fusion protein by cleavage with factor Xa. We demonstrated that the purified, recombinant SARS-CoV M^{pro} processes the peptide

H_2N-VSVNSTLQ↓SGLRKMA-COOH. This peptide, which represents the NH_2-terminal autoprocessing site of TGEV M^{pro} (cleavage site indicated by ↓; see Fig. 1B) and contains the sequence of our CMK inhibitor, is efficiently cleaved by SARS-CoV M^{pro} but not by its inactive catalytic-site mutant C145A (see Fig. 4).

Claim C: While peptidyl chloromethyl ketone inhibitors themselves are not useful as drugs because of their high reactivity and their sensitivity to cleavage by gastric and enteric proteinases, they are excellent substrate mimetics. With the CMK template structure at hand, we compared the binding mechanism to that seen in the distantly related picornavirus 3C proteinases ($3C^{pro}$). The latter enzymes have a chymotrypsin-related structure, similar to domains I and II of HCoV M^{pro}, although some of the secondary-structure elements are arranged differently, making structural alignment difficult (sequence identity <10%). Also, they completely lack a counterpart to domain III of coronavirus M^{pro}s. Nevertheless, the substrate specificity of picornavirus $3C^{pro}$s *(16,17)* for the P1', P1 and P4 sites is very similar to that of the coronavirus M^{pro}s *(8)*. As shown in Fig. 4, we found similar interactions between inhibitor and enzyme in case of the human rhinovirus (HRV) serotype 2 $3C^{pro}$ in complex with AG7088, an inhibitor carrying a vinylogous ethyl ester instead of a CMK group *(12)*. Only parts of the two structures can be spatially superimposed (r.m.s. deviation of 2.10 Å for 134 pairs of Cα positions out of the ~180 residues in domains I and II). Both inhibitors, the hexapeptidyl chloromethyl ketone and AG7088, bind to their respective target proteinases through formation of an antiparallel β-sheet with strand eII (Fig. 4). However, completely different segments of the polypeptide chain interact with the substrate analogs on the opposite site: residues 188 - 191 of the loop connecting domains II and III in M^{pro}, as opposed to the short β-strand 126 - 128 in HRV $3C^{pro}$. As a result, the architectures of the S2 subsites are entirely different

between the two enzymes; hence, the different specificities for the P2 residues of the substrates (Leu vs. Phe). The inhibitor AG7088 has a *p*-fluorophenylalanine side chain (*p*-fluorobenzyl) in this position. Based on molecular modeling, we believe that this side chain might be too long to fit into the S2 pocket of coronavirus M^{pro}, but an unmodified benzyl group would probably fit, as evidenced by Phe occuring in the P2 position of the COOH-terminal autocleavage site of the SARS coronavirus enzyme (deduced from the genomic sequence *(11)*). Apart from this difference, the superimposition of the two complexes (Fig. 4) suggests that the side chains of AG7088 binding to subsites S1 (lactone derivative of glutamine) and S4 (5-methyl-isoxazole-3-carbonyl) can be easily accommodated by the coronavirus M^{pro}. Thus, AG7088 could well serve as a starting point for modifications which should quickly lead to an efficient and bioavailable inhibitor for coronavirus main proteinases.

Since AG7088 is already clinically tested for treatment of the "common cold" (targeted at rhinovirus $3C^{pro}$), and since there are no cellular proteinases with which the inhibitors could interfere, prospects for developing broad-spectrum antiviral drugs on the basis of the structures presented here are good. Such drugs can be expected to be active against several viral proteinases exhibiting Gln↓(Ser,Ala,Gly) specificity, including the SARS coronavirus enzyme.

The structural information provided herein can be utilized to design or identify novel peptide drugs using, for example, a rational drug design (RDD) approach. Software applications typically utilized for such puposes include RIBBONS (Carson, M. (1997) Methods in Enzymology 277: 25), O (Jones, TA. *et al.* (1991) Acta Crystallogr A47:110), DINO (DINO: Visualizing Structural Biology (2001)

http://www.dino3d.org); and QUANTA, CHARMM, INSIGHT, SYBYL, MACROMODE, ICM, MOLMOL, RASMOL and GRASP (reviewed in Kraulis, J. (1991) Appl Crystallogr. 24:946). Additional information regarding RDD can be found in "Rational Drug Design" by Truhlar et al. (1999; Springer-Verlag New York, Incorporated).

The term "peptide" as used herein encompasses native peptides (either degradation products, synthetically synthesized peptides or recombinant peptides) and peptidomimetics (typically, synthetically synthesized peptides), as well as as peptoids and semipeptoids which are peptide analogs, which may have, for example, modifications rendering the peptides more stable while in a body or more capable of penetrating into cells. Such modifications include, but are not limited to N terminus modification, C terminus modification, peptide bond modification, including, but not limited to, CH2-NH, CH2-S, CH2-S=O, O=C-NH, CH2-O, CH2-CH2, S=C-NH, CH=CH or CF=CH, backbone modifications, and residue modification. Methods for preparing peptidomimetic compounds are well known in the art and are specified, for example, in Quantitative Drug Design, C.A. Ramsden Gd., Chapter 17.2, F. Choplin Pergamon Press (1992), which is incorporated by reference as if fully set forth herein. Further details in this respect are provided hereinunder.

Peptide bonds (-CO-NH-) within the peptide may be substituted, for example, by N-methylated bonds (-N(CH3)-CO-), ester bonds (-C(R)H-C-O-O-C(R)-N-), ketomethylen bonds (-CO-CH2-), α-aza bonds (-NH-N(R)-CO-), wherein R is any alkyl, e.g., methyl, carba bonds (-CH2-NH-), hydroxyethylene bonds (-CH(OH)-CH2-), thioamide bonds (-CS-NH-), olefinic double bonds (-CH=CH-), retro amide bonds (-NH-CO-), peptide derivatives (-N(R)-CH2-CO-), wherein R is the "normal" side chain, naturally presented on the carbon atom.

These modifications can occur at any of the bonds along the peptide chain and even at several (2-3) at the same time.

Natural aromatic amino acids, Trp, Tyr and Phe, may be substituted for synthetic non-natural acid such as TIC, naphthylelanine (Nol), ring-methylated derivatives of Phe, halogenated derivatives of Phe or o-methyl-Tyr.

In addition to the above, the peptides of the present invention may also include one or more modified amino acids or one or more non-amino acid monomers (e.g. fatty acids, complex carbohydrates etc).

The term "amino acid" or "amino acids" is understood to include the 20 naturally occurring amino acids; those amino acids often modified post-translationally *in vivo*, including, for example, hydroxyproline, phosphoserine and phosphothreonine; and other unusual amino acids including, but not limited to, 2-aminoadipic acid, hydroxylysine, isodesmosine, nor-valine, nor-leucine and ornithine. Furthermore, the term "amino acid" includes both D- and L-amino acids.

The peptides of the present invention are preferably utilized in a linear form, although it will be appreciated that in cases where cyclicization does not severely interfere with peptide characteristics, cyclic forms of the peptide can also be utilized.

The peptides of the present invention may be synthesized by any techniques that are known to those skilled in the art of peptide synthesis. For solid phase peptide synthesis, a summary of the many techniques may be found in J. M. Stewart and J. D. Young, Solid Phase Peptide Synthesis, W. H. Freeman Co. (San Francisco), 1963 and J. Meienhofer, Hormonal Proteins and Peptides, vol. 2, p. 46, Academic Press (New York), 1973. For classical solution synthesis see G. Schroder and K. Lupke, The Peptides, vol. 1, Academic Press (New York), 1965.

In general, these methods comprise the sequential addition of one or more amino acids or suitably protected amino acids to a growing peptide chain. Normally, either the amino or carboxyl group of the first amino acid is protected by a suitable protecting group. The protected or derivatized amino acid can then either be attached to an inert solid support or utilized in solution by adding the next amino acid in the sequence having the complimentary (amino or carboxyl) group suitably protected, under conditions suitable for forming the amide linkage. The protecting group is then removed from this newly added amino acid residue and the next amino acid (suitably protected) is then added, and so forth. After all the desired amino acids have been linked in the proper sequence, any remaining protecting groups (and any solid support) are removed sequentially or concurrently, to afford the final peptide compound. By simple modification of this general procedure, it is possible to add more than one amino acid at a time to a growing chain, for example, by coupling (under conditions which do not racemize

chiral centers) a protected tripeptide with a properly protected dipeptide to form, after deprotection, a pentapeptide and so forth. Further description of peptide synthesis is disclosed in U.S. Pat. No. 6,472,505.

A preferred method of preparing the peptide compounds of the present invention involves solid phase peptide synthesis.

Large scale peptide synthesis is described by Andersson Biopolymers 2000;55(3):227-50.

The peptides of the present invention can be provided to the subject *per se*, or as part of a pharmaceutical composition where it is mixed with a pharmaceutically acceptable carrier.

As used herein a "pharmaceutical composition" refers to a preparation of one or more of the active ingredients described herein with other chemical components such as physiologically suitable carriers and excipients. The purpose of a pharmaceutical composition is to facilitate administration of a compound to an organism.

Herein the term "active ingredient" refers to the preparation accountable for the biological effect.

Hereinafter, the phrases "physiologically acceptable carrier" and "pharmaceutically acceptable carrier" which may be interchangeably used refer to a carrier or a diluent that does not cause significant irritation to an organism and does not abrogate the biological activity and properties of the administered compound. An adjuvant is included under these phrases.

Since activity of peptides is directly correlated with a molecular weight thereof, measures are taken to conjugate the peptides of the present invention to high molecular weight carriers. Such high molecular weight carriers include, but are not limited to, polyalkylene glycol and polyethylene glycol (PEG), which are biocompatible polymers with a wide range of solubility in both organic and aqueous media (Mutter et al. (1979).

Alternatively, microparticles such as microcapsules or cationic lipids can serve as the pharmaceutically acceptable carriers of this aspect of the present invention.

As used herein, microparticles include liposomes, virosomes, microspheres and microcapsules formed of synthetic and/or natural polymers. Methods for making microcapsules and microspheres are known to the skilled in the art and

include solvent evaporation, solvent casting, spray drying and solvent extension. Examples of useful polymers which can be incorporated into various microparticles include polysaccharides, polyanhydrides, polyorthoesters, polyhydroxides and proteins and peptides.

Liposomes can be generated by methods well known in the art such as those reported by Kim et al., Biochim. Biophys. Acta, 728:339-348 (1983); Liu et al., Biochim. Biophys. Acta, 1104:95-101 (1992); and Lee et al., Biochim. Biophys. Acta, 1103:185-197 (1992); Wang et al., Biochem., 28:9508-9514 (1989). Alternatively, the peptide molecules of this aspect of the present invention can be incorporated within microparticles, or bound to the outside of the microparticles, either ionically or covalently.

As mentioned hereinabove the pharmaceutical compositions of this aspect of the present invention may further include excipients. The term "excipient", refers to an inert substance added to a pharmaceutical composition to further facilitate administration of an active ingredient. Examples, without limitation, of excipients include calcium carbonate, calcium phosphate, various sugars and types of starch, cellulose derivatives, gelatin, vegetable oils and polyethylene glycols.

Techniques for formulation and administration of drugs may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition, which is incorporated herein by reference.

Suitable routes of administration may, for example, include oral, rectal, transmucosal, especially transnasal, intestinal or parenteral delivery, including intramuscular, subcutaneous and intramedullary injections as well as intrathecal, direct intraventricular, intravenous, inrtaperitoneal, intranasal, or intraocular injections.

Alternately, one may administer a preparation in a local rather than systemic manner, for example, via injection of the preparation directly into a specific region of a patient's body.

Pharmaceutical compositions of the present invention may be manufactured by processes well known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes.

The peptide or peptides can be formulated into a composition in a neutral or salt form. Pharmaceutically acceptable salts, include the acid addition salts

(formed with the free amino groups of the peptide) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine, and the like

Pharmaceutical compositions for use in accordance with the present invention may be formulated in conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries, which facilitate processing of the active ingredients into preparations which, can be used pharmaceutically. Proper formulation is dependent upon the route of administration chosen.

For injection, the active ingredients of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hank's solution, Ringer's solution, or physiological salt buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for oral ingestion by a patient. Pharmacological preparations for oral use can be made using a solid excipient, optionally grinding the resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carbomethylcellulose; and/or physiologically acceptable polymers such as polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate.

Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used which may optionally contain gum

arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, titanium dioxide, lacquer solutions and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical compositions, which can be used orally, include push-fit capsules made of gelatin as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules may contain the active ingredients in admixture with filler such as lactose, binders such as starches, lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active ingredients may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for the chosen route of administration.

For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by nasal inhalation, the active ingredients for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from a pressurized pack or a nebulizer with the use of a suitable propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichloro-tetrafluoroethane or carbon dioxide. In the case of a pressurized aerosol, the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, e.g., gelatin for use in a dispenser may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

The preparations described herein may be formulated for parenteral administration, e.g., by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, e.g., in ampoules or in multidose containers with optionally, an added preservative. The compositions may be suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical compositions for parenteral administration include aqueous solutions of the active preparation in water-soluble form. Additionally, suspensions of the active ingredients may be prepared as appropriate oily or water based injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such

as sesame oil, or synthetic fatty acids esters such as ethyl oleate, triglycerides or liposomes. Aqueous injection suspensions may contain substances, which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the active ingredients to allow for the preparation of highly concentrated solutions.

Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, e.g., sterile, pyrogen-free water based solution, before use.

The preparation of the present invention may also be formulated in rectal compositions such as suppositories or retention enemas, using, e.g., conventional suppository bases such as cocoa butter or other glycerides.

Pharmaceutical compositions suitable for use in context of the present invention include compositions wherein the active ingredients are contained in an amount effective to achieve the intended purpose. More specifically, a therapeutically effective amount means an amount of active ingredients effective to prevent, alleviate or ameliorate symptoms of disease or prolong the survival of the subject being treated.

Determination of a therapeutically effective amount is well within the capability of those skilled in the art.

For any preparation used in the methods of the invention, the therapeutically effective amount or dose can be estimated initially from in vitro assays. For example, a dose can be formulated in animal models and such information can be used to more accurately determine useful doses in humans.

Toxicity and therapeutic efficacy of the active ingredients described herein can be determined by standard pharmaceutical procedures *in vitro*, in cell cultures or experimental animals. The data obtained from these in vitro and cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage may vary depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. (See e.g., Fingl, et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1).

Depending on the severity and responsiveness of the condition to be treated, dosing can be of a single or a plurality of administrations, with course of

treatment lasting from several days to several weeks or until cure is effected or diminution of the disease state is achieved.

The amount of a composition to be administered will, of course, be dependent on the subject being treated, the severity of the affliction, the manner of administration, the judgment of the prescribing physician, etc.

Compositions including the preparation of the present invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

Pharmaceutical compositions of the present invention may, if desired, be presented in a pack or dispenser device, such as an FDA approved kit, which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. The pack or dispenser may also be accommodated by a notice associated with the container in a form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals, which notice is reflective of approval by the agency of the form of the compositions or human or veterinary administration. Such notice, for example, may be of labeling approved by the U.S. Food and Drug Administration for prescription drugs or of an approved product insert.

EXAMPLES

Materials and Methods.

Protein expression and purification. Recombinant HCoV 229E $M^{pro}\Delta$(301-302) (residues 1 to 300; COOH-terminal residues 301 and 302 deleted) was expressed and purified essentially as described previously for the FIPV and full-length HCoV main proteinases *(13,15)*. Briefly, fusion proteins in which the HCoV pp1a/pp1ab amino acids 2966 to 3265 *(5)* had been fused to the *E. coli* maltose-binding protein (MBP), were expressed in *E.coli* TB1 cells (New England Biolabs). The fusion protein MBP-HCoV-$M^{pro}\Delta$(301-302) was purified by amylose-affinity chromatography and cleaved with factor Xa to release HCoV $M^{pro}\Delta$(301-302). Subsequently, the

recombinant proteinase was purified to homogeneity using phenyl Sepharose HP (Amersham Biosciences), Uno-Q (Bio-Rad Laboratories), and Superdex 75 (Amersham Biosciences) columns and concentrated to ≥15 mg/ml (Centricon-YM3, Millipore).

SARS-CoV M^{pro}Δ(305-306), which also had its two COOH-terminal residues deleted, was produced in an analogous way. As a control, a SARS-CoV M^{pro} mutant (SARS-CoV M^{pro}Δ(305-306)-C145A) was expressed and purified in an identical manner. In the latter, the active-site nucleophile, Cys^{145} (corresponding to Cys^{3385} of the pp1a/pp1ab polyprotein), was replaced by Ala. TGEV M^{pro} was expressed and purified as described *(10,15)*.

Preparation of selenomethionine-derivatized HCoV M^{pro}. To produce selenomethionine (SeMet)-substituted protein, the coding sequence of the MBP-HCoV-M^{pro}Δ(301-302) fusion protein was amplified by PCR and inserted into the unique *Nco*I site of pET-11d plasmid DNA (Novagen). The resulting plasmid, pET-HCoV-M^{pro}Δ(301-302), was used to transform the methionine auxotrophic 834(DE3) *E. coli* strain (Novagen), which was propagated in minimal medium containing 40 µg/ml seleno-L-methionine. The SeMet-substituted HCoV M^{pro}Δ(301-302) was purified as described above and concentrated to ≥7.1 mg/ml (Centricon-YM3, Millipore).

Dynamic light scattering. DLS experiments were performed using a DynaPro 801 device (Protein Solutions) with sample volumes of 15 µl.

Cleavage of a TGEV M^{pro} cleavage site by recombinant SARS-CoV M^{pro}. The peptide used in this assay was H_2N-VSVNSTLQSGLRKMA-COOH, which represents

the NH_2-terminal autocleavage site of TGEV M^{pro} *(9)* and corresponds to TGEV pp1a/pp1ab residues 2871 - 2885. The SARS-CoV $M^{pro}\Delta$(305-306) and M^{pro}(305-306)-C145A proteins (each at 0.5 μM) were incubated with 0.25 mM of the peptide for 45 min at 25°C in buffer consisting of 20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 1 mM EDTA, and 1 mM dithiothreitol. HPLC analysis of the cleavage reactions was done on a Delta Pak C_{18} column as described previously *(13)*.

Synthesis and purification of the hexapeptidyl chloromethyl ketone (Cbz-Val-Asn-Ser-Thr-Leu-Gln-CMK). Peptide synthesis was performed on an Applied Biosystems 433A peptide synthesizer using standard Fmoc-solid phase peptide synthesis protocols *(18)*. The reverse-phase HPLC chromatogram showed well-resolved peaks corresponding to the free NH_2-terminal peptide and the desired peptide carrying the Cbz group at the NH_2-terminal valine. The identity of the product was confirmed by mass spectrometry. Conversion of the free COOH-terminal of the purified, NH_2-protected peptide to the chloromethyl ketone functionality was performed as previously reported *(19)*. The product was then again purified by RP-HPLC and its identity confirmed by mass spectrometry.

Crystallization. Selenomethionine-HCoV $M^{pro}\Delta$(301-302) crystals were grown at 10 °C in hanging drops by equilibration of 7.1 mg/ml protein in 11 mM Tris-HCl (pH 8.0), 200 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 1% 1,6-hexanediol, and 10% polyethylene glycol 10,000 against 20% polyethylene glycol 10,000, 2% 1,6-hexanediol, 5 mM DTT, 12% dioxane and 100 mM HEPES, pH 8.5. Within about a week, fragile, plate-like crystals (~0.2x0.2x0.05 mm^3) were obtained. Crystals displayed space group $P2_1$ with unit cell dimensions a = 53.3 Å, b = 76.1 Å, c = 73.4 Å, β = 103.7°, and two proteinase monomers per asymmetric unit.

TGEV M^{pro} crystals were grown as described previously *(10)* and soaked for 16 h in a fivefold molar access of Cbz-Val-Asn-Ser-Thr-Leu-Gln-CMK, dissolved in a 1:1 mixture of dimethyl sulfoxide and acetonitrile. These crystals displayed space group $P2_1$ with unit cell dimensions a = 72.4 Å, b = 158.5 Å, c = 88.2 Å, β = 94.4°, and six proteinase molecules per asymmetric unit.

Collection of diffraction data. Using a Mar345 detector (X-ray Research), diffraction data from crystals of SeMet-HCoV $M^{pro}\Delta$(301-302) were collected at 100 K using synchrotron radiation at the XRD beamline of ELETTRA (Sincrotrone Trieste, Italy) at four different wavelengths around the selenium absorption edge (see Table 1). Due to the high concentration of polyethylene glycol in the mother liquor, these crystals did not require any cryoprotectant.

Crystals of TGEV M^{pro} that had been soaked with hexapeptidyl chloromethyl ketone inhibitor, were rinsed with mustard oil *(10)* before cryo-cooling in liquid nitrogen. A full diffraction data set was collected at 100 K, using the Joint IMB Jena/University of Hamburg/EMBL synchrotron beamline X13 at DESY (Hamburg, Germany) at a wavelength of 0.802 Å and equipped with a MarCCD detector (X-ray Research).

For both proteins, diffraction data were processed using the DENZO and SCALEPACK programs *(20)*. Diffraction data statistics are given in Table 1.

Structure solution. The anomalous signal from selenium in crystals of HCoV $M^{pro}\Delta$(301-302) was weak and did not provide sufficient phase information for solving the structure. Therefore, data collected at all four wavelengths were merged and used for structure elucidation by molecular replacement using AMoRe *(21)*, with a monomer of TGEV M^{pro} *(10)* as the search model (Table 2).

The structure of TGEV M^{pro} in complex with the hexapeptidyl chloromethyl ketone inhibitor was determined by difference Fourier methods. The maps showed density for all residues (except the benyzyloxycarbonyl (Cbz) protective group) of the inhibitor in the substrate-binding sites of monomers B and F. Density was weak at the Cβ atom of the P1 Gln residue, but the orientation of this side chain was still well defined due to the strong density for the carboxamide group. Density was also relatively weak for the side chains of the P5 and P6 residues of the inhibitor, indicating high mobility (particularly in the complex with monomer F). There was only little difference density near the S2 subsite in the substrate-binding clefts of the remaining four monomers, A, C, D, and E, indicating that these sites were occupied by 2-methyl-2,4-pentanediol (MPD) molecules from the crystallization medium, as in the free TGEV M^{pro} *(10)*.

Model building and refinement. Both the HCoV M^{pro} and TGEV M^{pro} CMK complex models were refined using CNS *(22)*. A random set of reflections containing 4% of the total data was excluded from the refinement for calculation of R_{free} *(23)*. Model building was carried out using the program 'O' *(24)*. σ_A-weighted maps *(25)* were used to avoid model bias. All residues of the HCoV $M^{pro}\Delta$(301-302) dimer were in unambiguous electron density. The final model comprises 600 amino-acid residues, 2 dioxane molecules and 221 water molecules. For the TGEV M^{pro} complex structure, all amino-acid residues in all six copies of the protein had well-defined electron density, with the exception of residues 301 and 302. The final model comprises 1799 amino-acid residues, 2 hexapeptidyl chloromethyl ketones, 4 MPD molecules, 27 sulfate ions, and 925 water molecules. Refinement statistics are summarized in Table 3.

Homology model building. InsightII (Molecular Simulations) was used to construct the three-dimensional model for SARS-CoV M^{pro} on the basis of the sequence alignment with HCoV M^{pro} and TGEV M^{pro}, and the crystal structures of these two enzymes. The model was energy-minimized in InsightII and inspected for steric consistency.

Analysis of the structural models. Overall geometric quality of the models was assessed using PROCHECK *(26)*. For HCoV M^{pro} and TGEV M^{pro}, respectively, 85.1% and 89.0% of the amino-acid residues were found in the most favored regions of the Ramachandran plot, and 15.5% and 10.5% were in additionally allowed regions. The corresponding numbers for the homology model of SARS-CoV M^{pro} were 87.1% and 11.3%. The agreement between structure-factor data and the atomic model was analyzed using SFCHECK *(27)*. Solvent accessibilities were calculated using the algorithm of Lee and Richards *(28)* as implemented in the program NACCESS (probe radius 1.4 Å). Molecular diagrams were drawn using the programs MOLSCRIPT *(29)*, PyMol *(30)*, and RASTER 3D *(31)*.

Table 1. Crystal Parameters and Statistics of Diffraction Data

Diffraction data statistics	HCoV M^{pro}	TGEV M^{pro}-CMK complex
Crystal Information		
Space group	$P2_1$	$P2_1$
Unit cell parameters (Å, °)	a = 53.3, b = 76.1, c = 73.4, β = 103.7	a = 72.4, b = 158.5, c = 88.2, β = 94.4
Estimated solvent content[a] (%)	44	51
Differaction data statistics		
X-ray source	Synchrotron radiation[b]	Synchrotron radiation[c]
Detector	Mar 345	MarCCD detector
No. of frames	600	720
Crystal oscillation (°)	1.0	0.5
Wavelength (Å)	0.980 (average)	0.802
Temperature (K)	100	100
Resolution (Å)[d]	25-2.60 (2.69-2.60)	50-2.37 Å (2.41-2.37)
Completeness (%)	98.9	99.8
R_{merge}(%)[d,e]	14.2 (41.2) [d]	8.0 (28.0)[d]
R_{rim} (%)[d,f]	14.2 (43.9)[d]	2.2 (8.5)[d]
R_{pim} (%)[d,g]	3.0 (13.0)[d]	0.058 (22.0)[d]
Redundancy	12.3	7.1
$I/\sigma(I)$	9.1	9.9
Mosaicity (°)	1.80	0.49

No. of reflections measured	216,984	569,126
Unique reflections	17,533	79,667

[a] Solvent content estimated according to *(32)*.

[b] X-ray diffraction beamline at ELETTRA, Trieste, equipped with a Mar345 detector

[c] Joint IMB Jena/University of Hamburg/EMBL synchrotron beamline X13 at Deutsches Elektronen-Synchrotron (DESY), Hamburg, equipped with a MarCCD detector

[d] Highest resolution bin in parentheses

[e] $R_{merge}=100 \times \Sigma_i\Sigma_{hkl}|I_i - \langle I\rangle| / \Sigma_i\Sigma_{hkl}I_i$, where I_i is the observed intensity and $\langle I\rangle$ is the average intensity from multiple measurements

[f] $R_{rim} = 100 \times \Sigma_l (N/N-1)^{1/2}\Sigma_{hkl}|I_i-\langle I\rangle|/\Sigma_i\Sigma_{hkl}I_i$ where N is the number of times a given reflection has been measured. This quality indicator corresponds to an R_{sym} that is indenpendent of the redundancy of the measurements *(33)*.

[g] $R_{pim} = 100 \times \Sigma_i (1/N-1)^{1/2}\Sigma_{hkl}|I_i-\langle I\rangle|/\Sigma_i\Sigma_{hkl}I_i$. This factor provides information about the average precision of the data *(33)*.

Table 2: Structure solution by molecular replacement: HCoV M^{pro}

Resolution range	10.0 - 4.0 Å
Rotation and translation function (1st monomer)	
Best solution	□ =21.64°, β = 59.58°, γ = 256.95° tx = 0.483, ty = 0.000, tz = 0.250 Å
Correlation coefficient	0.217
R-factor	51.9%
Rotation and translation function (2nd monomer)	
Best solution	□ = 319.92°, β = 79.38°, γ = 5.39° tx = 0.054, ty = 0.481, tz = 0.785 Å
Correlation coefficient	0.213
R-factor	52.1%
Refinement of combined solution	
Monomer 1	□ = 21.80°, β = 60.40°, γ = 257.02° tx =0.478, ty = -0.002, tz = 0.250 Å
Monomer 2	□ = 320.45°, β = 79.89°, γ = 5.89° tx = 0.057, ty = 0.482, tz = 0.784 Å
Correlation coefficient	0.30
R-factor	48.8%

Table 3: Phasing and refinement statistics, and model quality

Phasing	**HCoV M^{pro}**	**TGEV M^{pro}-C-Mk complex**
Refinement		
Resolution range (Å)	25 – 2.6	50 – 2.37
R factor[a]	0.219	19.1
R_{free}	0.283	23.5
No. of non-hydrogen atoms (average B value ($Å^2$))		
Protein	4594 (28.12)	13,819 (43.0)
Water	221 (24.9)	925 (51.3)
MPD	-	32 (78.6)
Sulfate	-	135 (59.8)
Dioxane	12 (58.39)	-
Substrate-analog inhibitor	-	92 (71.0)
Bonds (Å)	0.012	0.006
Angles (°)	1.5	1.3

[a]R-factor = Σ ($|F_o|$ - k$|F_c|$) / Σ $|F_o|$

REFERENCES

1. S. H. Myint, in *The Coronavirdae,* S. G. Siddell, Ed. (Plenum Press, New York, 1995), pp. 389.

2. C. Drosten *et al., Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome* (http://content.nejm.org/cgi/content/abstract/NEJMoa030747v2) *N. Engl. J. Med.,* in the press (2003).

3. T. G. Ksiazek *et al., A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome* (http://content.nejm.org/cgi/content/abstract/NEJMoa030781v2) *N. Engl. J. Med.*, in the press (2003) .

4. N. Lee *et al., A Major Outbreak of Severe Acute Respiratory Syndrome in Hong Kong* (http://content.nejm.org/cgi/content/abstract/NEJMoa030685v1) *N. Engl. J. Med.,* in the press (2003).

5. J. Herold, T. Raabe, B. Schelle-Prinz, S. G. Siddell, *Virology* **195**, 680 (1993).

6. V. Thiel, J. Herold, B. Schelle, S. G. Siddell, *J. Gen. Virol.* **75**, 6676 (2001).

7. J. Ziebuhr, J. Herold, S. G. Siddell, *J. Virol.* **69**, 4331 (1995).

8. J. Ziebuhr, E. J. Snijder, A. E. Gorbalenya, *J. Virol.* **81**, 853 (2000).

9. A. Hegyi, J. Ziebuhr, *J. Gen. Virol.* **83**, 595 (2002).

10. K. Anand *et al.*, *EMBO J.* **21**, 3213 (2002).

11. M. Marra *et al.*, http://www.bcgsc.ca/bioinfo/SARS/

12. D. A. Matthews *et al., Proc. Natl. Acad. Sci. USA* **96**, 11000 (1999).

13. J. Ziebuhr, G. Heusipp, S. G. Siddell, *J. Virol.* **71**, 3992 (1997).

14. S. G. Siddell, in *The Coronaviridae*, S. G. Siddell, Ed. (Plenum Press, New York, 1995), p. 1

15. A. Hegyi, A. Friebe, A. E. Gorbalenya, J. Ziebuhr, *J. Gen. Virol.* **83**, 581 (2002).

16. H. G. Kräusslich, E. Wimmer, *Annu. Rev. Biochem.* **57**, 701 (1988).

17. M. D. Ryan, M. Flint, *J. Gen. Virol.* 78, 699 (1997).

18. G. B. Fields, R. L. Noble, *Int. J. Pept. Prot. Res.* **35**, 161 (1990).

19. A. Krantz, L. J. Copp, P. J. Coles, R. A. Smith, S. B. Heard, *Biochemistry* **30**, 4678 (1991).

20. Z. Otwinowski, W. Minor *Methods Enzymol.* **276**, 307 (1997).

21. J. Navaza, *Acta Crystallogr.* **A50**, 157 (1994).

22. A. T. Brünger *et al., Acta Crystallogr.* **D54**, 905 (1998).

23. A. T. Brünger, *Nature* **355**, 472 (1992).

24. T. A Jones, S. Cowan, J-Y. Zou, M. Kjeldgaard, *Acta Crystallogr.* **A47**, 110 (1991).

25. R. J. Read, *Acta Crystallogr.* **A42**, 140 (1986).

26. R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Crystallogr.* **26**, 283 (1993).

27. A. A. Vaguine, J. Richelle, S. J. Wodak, *Acta Crystallogr.* **D55**, 191 (1999).

28. B. Lee, F. M. Richards, *J. Mol. Biol.* **55**, 379 (1971).

29. P. J. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991).

30. W. L. DeLano, *The PyMOL Molecular Graphics System.* DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org/ (2002).

31. E. A. Merritt, D. J. Bacon, *Meth. Enzymol.* **277**, 505 (1997).

32. B. W. Matthews, *J. Mol. Biol.* **33**, 491 (1968).

33. M. S. Weiss, R. Hilgenfeld, *J. Appl. Crystallogr.* **30**, 203 (1997).

Description of the Annex content

The Annex enclosed herewith contains the print of the following 3 text files:

PDB file1

PDB file2

PDB file3

WHAT IS CLAIMED IS:

1. A composition-of-matter comprising a crystallized complex including a porcine transmissible gastroenteritis (corona)virus (TGEV) main proteinase and an inhibitor thereof.

2. A computing platform for generating a 3D atomic structure model of at least a portion of a complex including at least a porcine transmissible gastroenteritis (corona)virus (TGEV) main proteinase, the computing platform comprising:

(a) a data-storage device storing data comprising a set of structure coordinates defining at least a portion of a 3D atomic structure of the complex; and

(b) a processing unit being for generating the 3D atomic structure model from said data stored in said data-storage device.

3. A computer readable medium comprising, in a retrievable format, data including a set of structure coordinates defining at least a portion of a 3D atomic structure of a porcine transmissible gastroenteritis (corona)virus (TGEV) main proteinase.

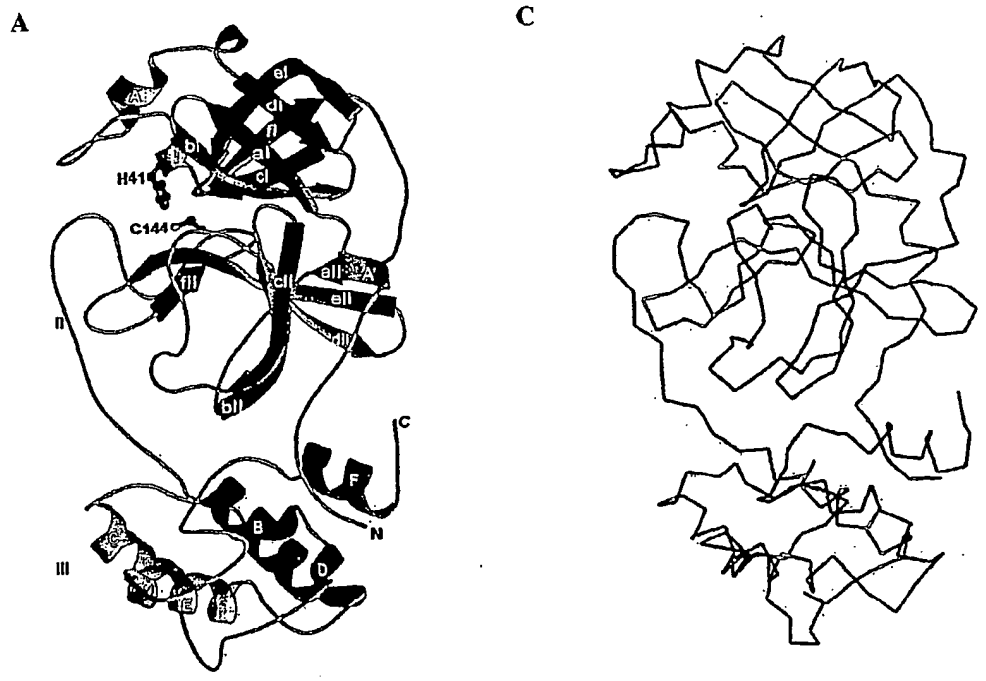
4. A computer generated model representing at least a portion of a 3D atomic structure of a porcine transmissible gastroenteritis (corona)virus (TGEV) main proteinase.

5. A method of treating SARS in an individual comprising administering to the individual a therapeutically effective amount of a peptide inhibitor capable of binding corona virus main proteinase, thereby treating SARS in the individual.

6. A method of designing a SARS inhibitor comprising utilizing the following three sets of atomic coordinates:(1) crystal structure of human coronavirus 229E (HCoV) main proteinase (PDB file no. 1) (2) model structure of SARS-associated coronavirus (SARS-CoV) main proteinase, based on the crystal structure in claim A1 (PDB file no. 2) and (3) crystal structure of transmissible gastroenteritis virus (TGEV) main proteinase in complex with a hexapeptidyl chloromethylketone inhibitor (PDB file no. 3) in modeling SARSprotease inhibition, thereby designing a SARS inhibitor.

7. A method to produce enzymatically active SARS-CoV main proteinase and modifications (mutants) thereof

8. The use of Michael acceptor compounds having α,β-unsaturated carbonyl groups as inhibitors for coronavirus main proteinases



B

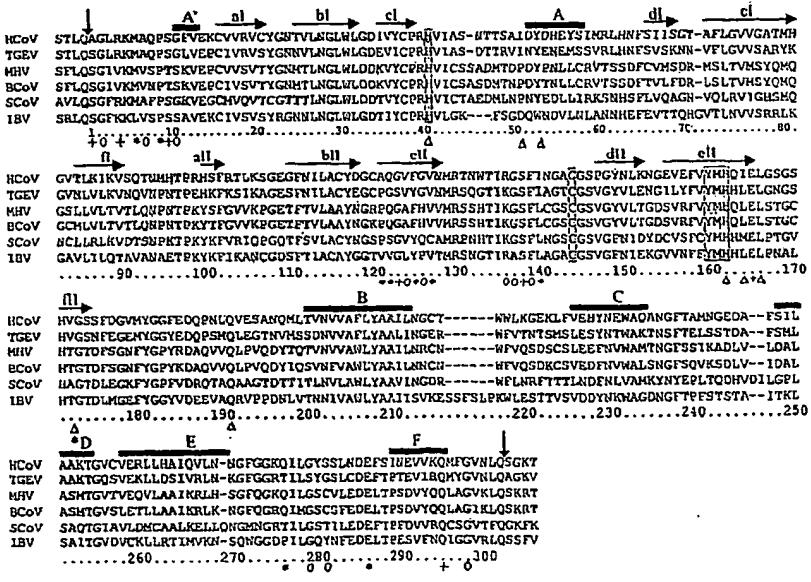
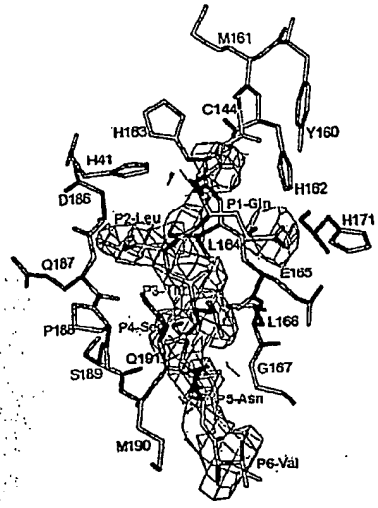


Figure 1

Figure 2

A


M161
C144
H163
Y160
H41
H162
D186
P1-Gln
H171
P2-Leu
L164
Q187
E165
P3-Thr
P188
P4-Ser
L166
Q191
G167
S189
P5-Asn
M190
P6-Val

B

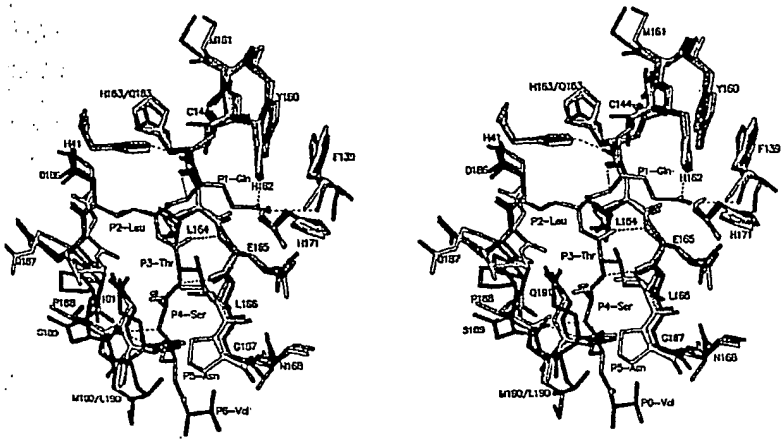


Figure 3

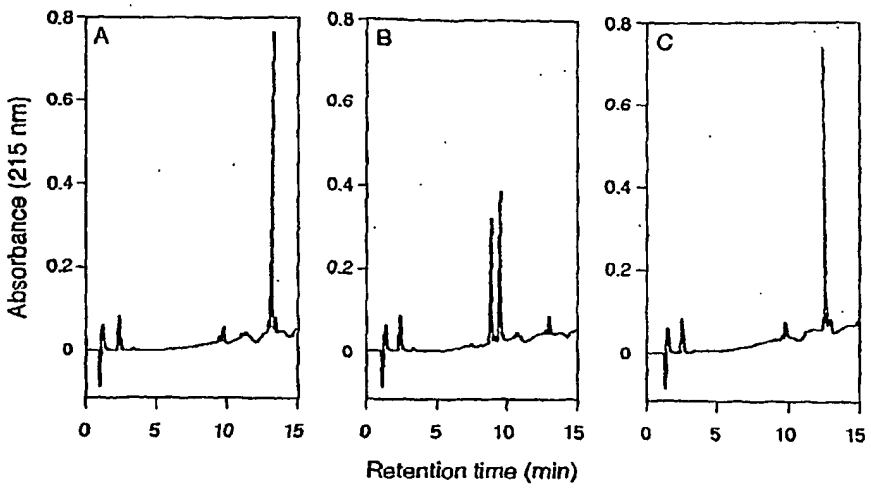


Figure 4

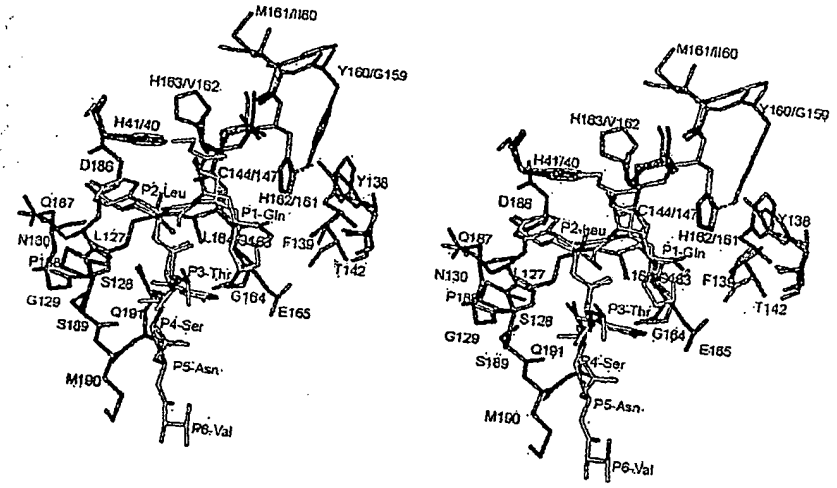

M161/I160
Y160/G159
H163/V162
H41/40
D186
C144/147
Y138
Q187
P2-Leu
H162/161
P1-Gln
N130
L127
F139
T142
P3-Thr
S128
G164
E165
G129
Q191
P4-Ser
S189
P5-Asn
M190
P6-Val

Figure 5

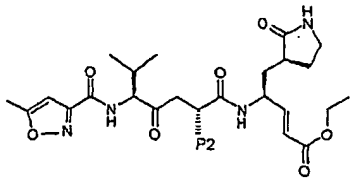


Figure 6

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP2004/005109

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N9/50 A61P31/14 G01N33/50

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ANAND KANCHAN ET AL: "Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain" EMBO (EUROPEAN MOLECULAR BIOLOGY ORGANIZATION) JOURNAL, vol. 21, no. 13, 1 July 2002 (2002-07-01), pages 3213-3224, XP002299685 ISSN: 0261-4189 in particular see page 3214 the structure determination of the TGEV MPro molecule and material and methods page 3222 the whole document ----- -/--	1,5-8

[X] Further documents are listed in the continuation of box C. [ ] Patent family members are listed in annex.

° Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
7 October 2004	22/10/2004

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Vix, O

Form PCT/ISA/210 (second sheet) (January 2004)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	ANAND KANCHAN ET AL: "Coronavirus main proteinase (3CLpro) structure: Basis for design of anti-SARS drugs" SCIENCE, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE,, US, vol. 300, no. 5626, 13 June 2003 (2003-06-13), pages 1763-1767, XP002291115 ISSN: 0036-8075 the whole document	1,5-8
P,X	CHOU K-C ET AL: "Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS" BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, ACADEMIC PRESS INC. ORLANDO, FL, US, vol. 308, no. 1, 15 August 2003 (2003-08-15), pages 148-151, XP004441432 ISSN: 0006-291X the whole document	1,5-8
P,X	XUE WU ZHANG ET AL: "EXPLORING THE BINDING MECHANISM OF THE MAIN PROTEINASE IN SARS-ASSOCIATED CORONAVIRUS AND ITS IMPLICATION TO ANTI-SARS DRUG DESIGN" BIOORGANIC & MEDICINAL CHEMISTRY, ELSEVIER SCIENCE LTD, GB, vol. 12, no. 9, 1 May 2004 (2004-05-01), pages 2219-2223, XP001202604 ISSN: 0968-0896 the whole document	1,5-8
P,X	YANG HAITAO ET AL: "The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, vol. 100, no. 23, 11 November 2003 (2003-11-11), pages 13190-13195, XP002299686 ISSN: 0027-8424 the whole document	1,5-8

-/--

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	JENWITHEESUK EKACHAI ET AL: "Identifying inhibitors of the SARS coronavirus proteinase." BIOORGANIC & MEDICINAL CHEMISTRY LETTERS, vol. 13, no. 22, 17 November 2003 (2003-11-17), pages 3989-3992, XP002299687 ISSN: 0960-894X the whole document	5-8
A	KUHN P ET AL: "The genesis of high-throughput structure-based drug discovery using protein crystallography" CURRENT OPINION IN CHEMICAL BIOLOGY, CURRENT BIOLOGY LTD, LONDON, GB, vol. 6, no. 5, October 2002 (2002-10), pages 704-710, XP002287759 ISSN: 1367-5931 the whole document	6

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box II.1

Claims Nos.: 2-4

Claims 2-4-Presentation of information
Concerning claims 2-4 applicant's attention is drawn to Rule 39.1(v) PCT. The subject-matter of claims 2-4 refers to the presentation of structure data (3D atomic structure model of TGEV main protease) and is not regarded as patentable invention within the meaning of Rule 67 (v) PCT since it relates to a presentation of information (protein model structure coordinates or computer generated model) as a coordinate listings (or structure factor data) and their use, or information stored on a computer (e.g. data storage device storing data on a computer) or computer readable media. Thus, the above mentioned claims will not be searched.

Other remark:
Although claim 5 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2004/005109

Box II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.: 2-4
because they relate to subject matter not required to be searched by this Authority, namely:

see FURTHER INFORMATION sheet PCT/ISA/210

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.